

# ATLAS Distributed Analysis: Current roadmap

## Closeout session Distributed Analysis Review

David Adams – DIAL/PPDG/BNL

Dietrich Liko – ARDA/EGEE/CERN

Karl Harrison – GANGA/GridPP/Cambridge

Alvin Tan – GridPP/Birmingham

July 13, 2005



D. Adams, D. Liko,  
K...Harrison, C. L. Tan

# Contents

Goals

Model

Deployment

Production

Data management

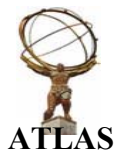
Current status

Short term plans

Components

Contributors

Milestones



D. Adams, D. Liko,  
K...Harrison, C. L. Tan

ADA Review

Current roadmap

July 13, 2005 2

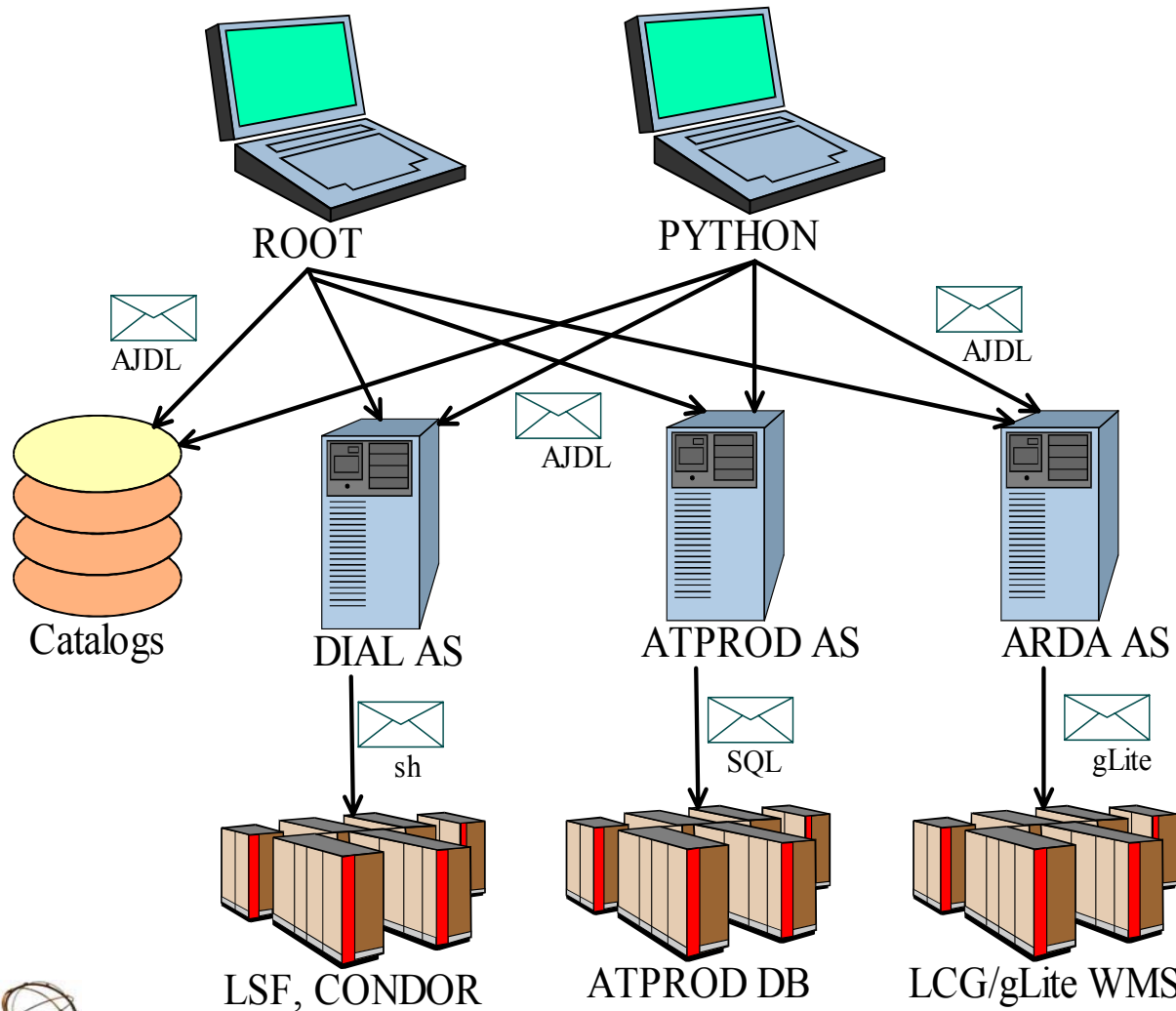
# Goals

## Goals of the ADA system

- Provide processing access to all ATLAS event data
  - Database group has responsibility for placement and cataloging of production data and some analysis results
- Enable effective use of all ATLAS computing resources
  - Including private or opportunistic
- Fair allocation between users, physics groups, ...
- Easy-to-use interface
- Full provenance tracking
  - Where did this data (event or analysis) come from?



# Model



GUI and  
command line  
clients

High level services  
for cataloging and  
job submission and  
monitoring

Workload  
management  
systems



D. Adams, D. Liko,  
K...Harrison, C. L. Tan

ADA Review

Current roadmap

July 13, 2005 4

# Model (cont)

## AJDL – Abstract Job Definition Language

- Dataset used to describe input and output data
  - Event data or analysis (ntuple, histograms)
- Transformation describes action to take on the data
  - Application scripts used to run job to build task or process data
  - Task carries user parameters or code
    - > E.g. atlas release, job options, and/or algorithm code
- Job is an instance of a transformation applied to a dataset
  - Job splitting accomplished by splitting input dataset and merging the resulting output datasets

## Catalogs

- Repositories to hold instances of application, task and dataset
- Selection catalogs associate names and metadata with instances of these objects



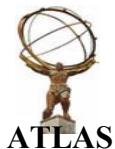
# Model (cont)

## Analysis services

- Receive job request (transformation plus dataset)
- Process job
  - Build task or locate built task
  - Split input dataset
  - Submit subjob for each subdataset
  - Merge results from subjobs
- Provide means to check job status
- Provide means to kill job

## Clients

- Provide means to access catalogs and analysis services
- Easy-to-use
- Integrated with ATLAS analysis environments
  - Root, python



# Deployment

## Hierarchy of analysis services

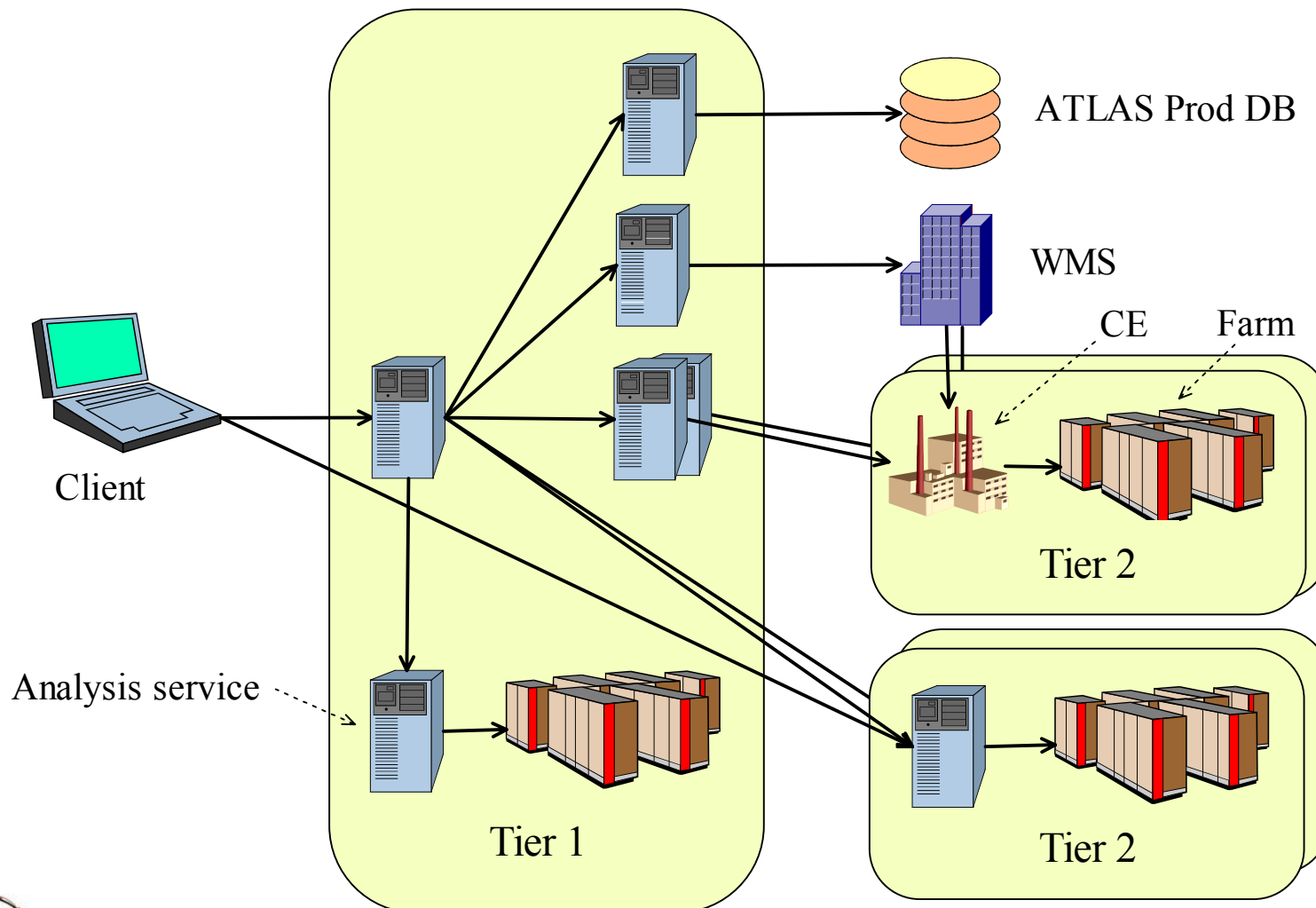
- Distribute the load of job submission, monitoring and merging
- Operate services close to resources
  - Service provides common interface to monitor performance and report status of underlying resources
    - > submission rate and efficiency, latencies, data throughput
- Figure shows possible ATLAS deployment strategy
  - Tier 1 sites maintain most of the analysis services
  - Tier 2 has option to provide analysis service or grid CE

## Interactive (responsive) services add requirements

- High job rates ( $> 1$  Hz)
- Low submit-to-result job latencies ( $< 1$  minute)
- High data input rate from SE to farm ( $> 10$  MB/job)



# Deployment (cont)





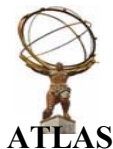
# Production

Distributed analysis differs from production in many ways

- Many users *vs.* one
- Often want interactive response
- Often need prompt access to results
- Large number of short-running jobs
- Many results are temporary or short-lived
- Users are not grid or computing experts
- Typically need to provide user code

## Integration

- Current production system is far from handling the above
- Sufficiently different requirements to justify separate projects
- However, there is need for coordination
  - Production system might handle a large fraction of analysis jobs
  - Common accounting system

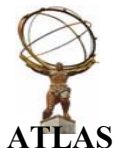


# Data management

Distributed analysis has special data management needs

- Often want interactive response for locating and moving files
- Many types of objects
  - Files, datasets, transformations, jobs
  - Need for users to *easily* control lifetimes
- Large number of objects with short (but finite) lifetimes
- Many people owning and managing these objects
- May be multiple owners of some
- Provenance tracking
  - Record the transformation and input dataset for each output dataset
  - Full chain including the same for the input dataset
  - Implies provenance table also owns its entries

Integration with DMS to address all the above



# Current status

## Datasets

- Descriptions for Rome AOD
- Files at BNL and (smaller datasets) at CERN

## Transformations

- Many ATLAS-specific transformations have been defined
  - All run athena
  - Characterized by task data
    - > Atlasopt: ATLAS release and job options
    - > Aodhisto: atlasopt plus code to build in UserAnalysis package
    - > Atlasdev: atlasopt plus local development directory
    - > Atlasdev-src: same as atlasdev except development area is tarred up and will be rebuilt if platform changes



# Current status (cont)

## Analysis services

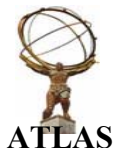
- BNL LSF-based service is almost always available
  - Jobs submitted to local no-wait LSF queue
  - Plot show recent performance measurement
- CERN intermittently provides LCG/gLite service
  - More reliable once gLite is available?

## Catalogs

- BNL maintains MySQL catalogs
  - Dataset, application and task repositories and selection catalogs

## Clients

- Root-based client is described in “user guide”
- Python client is similar
- GUI available for job submission and monitoring
- Web-based service monitor and catalog browsers available



## ATLAS Distributed Analysis Service Monitor

URL	Site	Status	Description
adial01.usatlas.bnl.gov:20011	BNL	Valid	ATLAS interactive analysis service [1.20]
adial01.usatlas.bnl.gov:20001	BNL	Valid	ATLAS unique ID service [1.20]

Last update: 03:06:41 07/13/05 EDT

Next update in about 5 minutes.



D. Adams, D. Liko,  
K...Harrison, C. L. Tan

ADA Review

Current roadmap

July 13, 2005 13

ADA Job Builder v1.0.3

Job Builder | Job Monitoring | ADA Settings

UId:

Name:





SQL Query:

Show first  results of the query.

SQL Query Builder

Condition	Field	Operator	Value
AND	uid	=	<input type="text"/>

Selection Contents

 build\_task
  readme.txt
  release\_notes.txt
  run

Application | Task | Dataset | Preferences

Submit  
Load  
Save  
Quit

Checklist

- ☐ Application
- ☐ Task
- ☐ Dataset
- ☐ Preferences



ADA Job Builder v1.0.3

Job Builder | **Job Monitoring** | ADA Settings

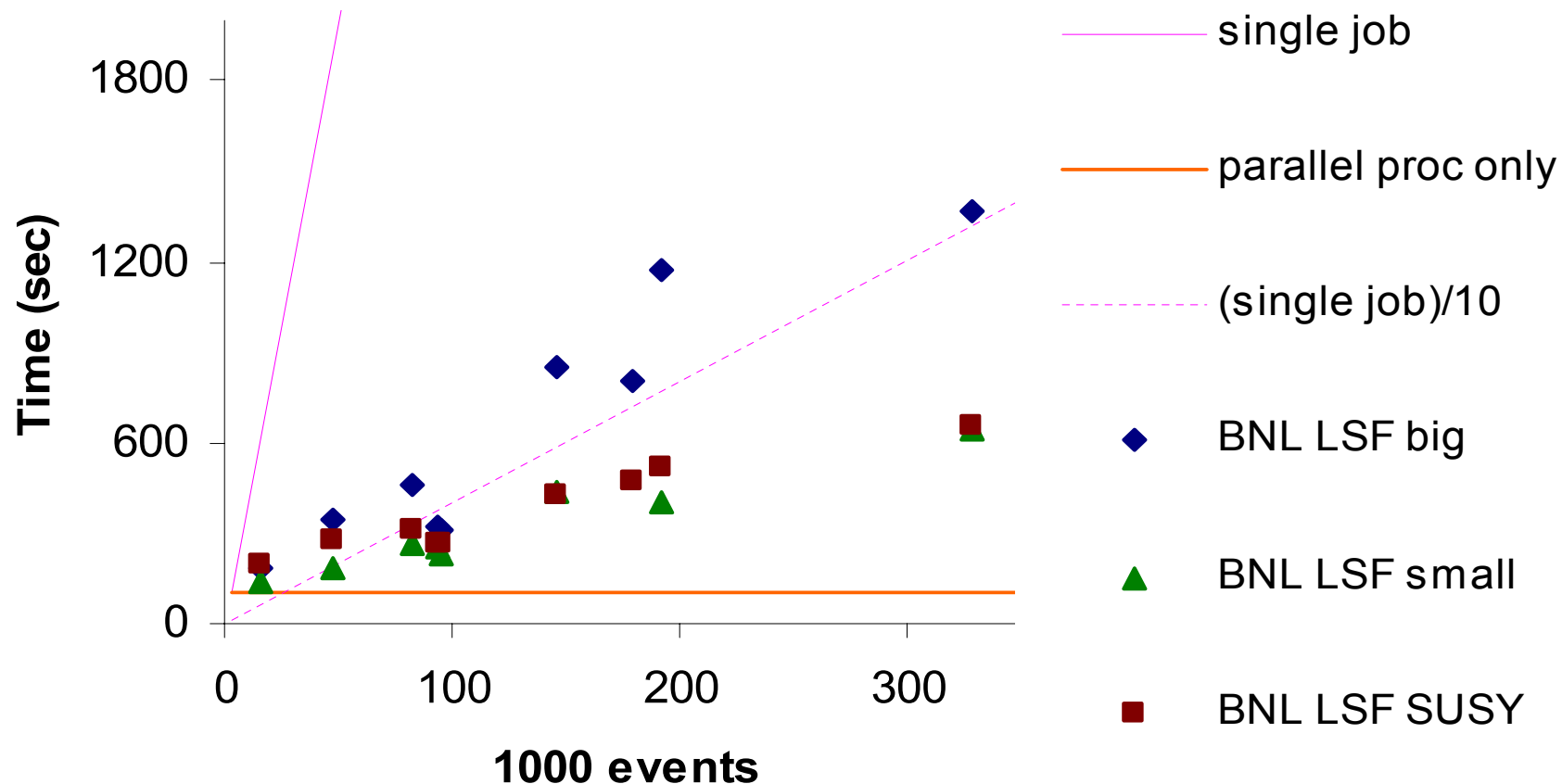
Monitoring Server Status: **OK** [Connect](#)

	id	status	has_result	duration	owner
1	10501-48386	DONE	Yes	0 hrs: 6 mins: 19 secs	/DC=org/DC=doegrids/OU=People/CN=David Adams 4071
2	10501-48467	DONE	Yes	0 hrs: 6 mins: 25 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
3	10501-48548	DONE	Yes	0 hrs: 9 mins: 43 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
4	10501-48629	DONE	Yes	0 hrs: 6 mins: 24 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
5	10501-48710	DONE	Yes	0 hrs: 7 mins: 28 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
6	10501-48791	DONE	Yes	0 hrs: 6 mins: 31 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
7	10501-48872	DONE	Yes	0 hrs: 4 mins: 35 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
8	10501-48877	DONE	Yes	0 hrs: 5 mins: 50 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
9	10501-48882	DONE	Yes	0 hrs: 5 mins: 30 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
<b>10</b>	<b>10501-48887</b>	<b>DONE</b>	<b>Yes</b>	<b>0 hrs: 5 mins: 37 secs</b>	<b>/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136</b>
11	10501-48892	DONE	Yes	0 hrs: 5 mins: 36 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
12	10501-48897	DONE	Yes	0 hrs: 1 mins: 10 secs	/C=UK/O=eScience/OU=Birmingham/L=ParticlePhysics/CN=
13	10501-48908	DONE	Yes	0 hrs: 0 mins: 56 secs	/C=UK/O=eScience/OU=Birmingham/L=ParticlePhysics/CN=
14	10501-48919	FAILED	No	0 hrs: 0 mins: 18 secs	/C=UK/O=eScience/OU=Birmingham/L=ParticlePhysics/CN=
15	10501-48930	DONE	Yes	0 hrs: 4 mins: 31 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
16	10501-48935	DONE	Yes	0 hrs: 4 mins: 22 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
17	10501-48940	DONE	Yes	0 hrs: 5 mins: 8 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
18	10501-48949	DONE	Yes	0 hrs: 4 mins: 36 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
19	10501-48958	DONE	Yes	0 hrs: 4 mins: 32 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
20	10501-48967	DONE	Yes	0 hrs: 4 mins: 19 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
21	10501-48976	DONE	Yes	0 hrs: 3 mins: 50 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
22	10501-48985	DONE	Yes	0 hrs: 3 mins: 47 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
23	10501-48994	DONE	Yes	0 hrs: 3 mins: 51 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136
24	10501-49003	DONE	Yes	0 hrs: 4 mins: 20 secs	/DC=org/DC=doegrids/OU=People/CN=Frank E. Paige 136

[Options](#) [Force Refresh](#)



## ADA AOD processing time using DIAL 1.20 29jun05



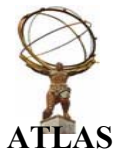
D. Adams, D. Liko,  
K...Harrison, C. L. Tan



# Short term plans

## Datasets

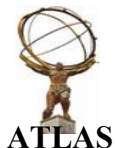
- Add command to create dataset with full structure
  - In current development release
  - Add transformation to distribute this?
- Add Rome ESD
- Dataset to describe event collection
- Connect to DMS as discussed earlier
  - Rome data not yet included in DMS



# Short term plans (cont)

## Analysis services

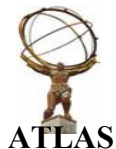
- Add persistence
  - Record input dataset, transformation, job and output dataset in repositories
  - Resolve memory leaks
- Improve reliability
  - Last release has already proved stable
- ATPROD service
- LCG/gLite service
- Forwarding service
  - Based on locally supplied algorithm using resource and data availability
- Add means for service to report data and resource availability
- Deploy at more sites
- Add means to monitor service performance



# Short term plans (cont)

## Catalogs

- Add local repository caches
  - Local catalog that retrieves and stores data from central repository
    - > Objects are immutable so no coherency problems
- AMI or other ATLAS DB deployment of primary catalogs



# Components

## User Interface [1.0]

- Root - DIAL
- Python - GANGA
- Command line - DIAL
- Service monitor - GANGA
- Job GUI - GANGA

## Services [1.0]

- Infrastructure - DIAL
- LSF - DIAL
- Condor - DIAL
- EGEE/LCG - ARDA
- ATPROD - ATPROD

## Service deployment [1.0]

- BNL - DIAL
- CERN - ARDA
- Other sites - DIAL, ARDA, ADA

## Data [0.25]

- ATLAS Dataset classes - DIAL
- Rome AOD,ESD datasets - DIAL

## Transformations [0.25]

- Analysis - ADA
- Production - ADA, ATPROD

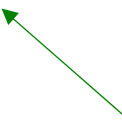
## Catalogs [0.5] - DIAL, AMI

## Software management [0.2] - ADA

## Accounting and allocation [1.0]

## Integration and testing [0.5] - ALL

TOTAL [5.7]



*Component*



*Estimated future FTE's*



*Current contributors*



# Contributors

## GANGA

- Karl Harrison, Alvin Tan

## DIAL

- David Adams, Wensheng Deng, Tadashi Maeno, Vinay Sambamurthy, Nagesh Chetan, Chitra Kannan

## ARDA

- Dietrich Liko

## ATPROD

- Frederic Brochu, Alessandro De Salvo

## AMI

- Solveig Albrand, Jerome Fulachier

## ADA

- (plus those above), Farida Fassi, Christian Haeberli, Hong Ma, Grigori Rybkine



# Milestones

2007

- Access to all MC and reconstructed data as it is placed at sites
  - Interactive, WMS and ATPROD analysis services

Fall 2005

- Most Rome AOD and ESD available for ADA analysis
- Analysis service enabling use of ATLAS production system
- Large scale user analysis and production
  - LCG/gLite WMS
  - Multiple LCG sites
  - Job placement based on dataset placement
  - Job requests handled by analysis service at CERN
- Interactive response for short-running jobs
  - Distributed analysis services or interactive CE's
  - Multiple sites; choice based on dataset placement
  - Job requests handled by analysis service at BNL
- Expect Rome data to be in DMS

